buildings

MDPI

Article

Achieving On-Site Trustworthy AI Implementation in the Construction Industry: A Framework Across the AI Lifecycle

Lichao Yang [1], Gavin Allen [2], Zichao Zhang [1] and Yifan Zhao [1,*]

[1] Faculty of Engineering and Applied Science, Cranfield University, Cranfield MK43 0AL, UK; lichao.yang@cranfield.ac.uk (L.Y.); zichao.zhang.960@cranfield.ac.uk (Z.Z.)
[2] Glideology, Malvern WR14 4PA, UK; gavin.allen@glideology.co.uk
* Correspondence: yifan.zhao@cranfield.ac.uk

Abstract: In recent years, the application of artificial intelligence (AI) technology in the construction industry has rapidly emerged, particularly in areas such as site monitoring and project management. This technology has demonstrated its great potential in enhancing safety and productivity in construction. However, concerns regarding the technical maturity and reliability, safety, and privacy implications have led to a lack of trust in AI among stakeholders and end users in the construction industry, which slows the intelligent transformation of the industry, particularly for on-site AI implementation. This paper reviews frameworks for AI system design across various sectors and government regulations and requirements for achieving trustworthy and responsible AI. The principles for the AI system design are then determined. Furthermore, a lifecycle design framework specifically tailored for AI systems deployed in the construction industry is proposed. This framework addresses six key phases, including planning, data collection, algorithm development, deployment, maintenance, and archiving, and clarifies the design principles and development priorities needed for each phase to enhance AI system trustworthiness and acceptance. This framework provides design guidance for the implementation of AI in the construction industry, particularly for on-site applications, aiming to facilitate the intelligent transformation of the construction industry.

Keywords: artificial intelligence; trustworthy and responsible AI; digital construction

Academic Editor: Jorge Lopes

Received: 5 November 2024
Revised: 9 December 2024
Accepted: 21 December 2024
Published: 25 December 2024

Citation: Yang, L.; Allen, G.; Zhang, Z.; Zhao, Y. Achieving On-Site Trustworthy AI Implementation in the Construction Industry: A Framework Across the AI Lifecycle. *Buildings* **2025**, *15*, 21. https://doi.org/10.3390/buildings15010021

1. Introduction

The rapid population increase will significantly heighten the demand for infrastructure, presenting substantial challenges to the current productivity levels in the construction industry [1]. Meanwhile, a report published by the United Nations Environment Programme in 2023 revealed that the buildings and construction sector is responsible for 37% of global greenhouse gas emissions [2], making it one of the largest contributors to climate change. To achieve the net zero emissions target by 2050, optimising the current production pattern to reduce emissions and increase productivity becomes the most important challenge facing the construction industry. In recent years, the use of artificial intelligence (AI) and other digital technologies has been recognised as an effective way to improve productivity due to their capability to reduce costs and improve management efficiency. These technologies have driven digitalisation and automation across various industries. However, the construction industry's adoption of digital technologies remains low due to several factors [3,4]. First, this sector typically operates on low-profit margins, making it cautious about investing in new technologies. Second, there is a significant shortage of specialised knowledge and skilled technicians to implement these technologies effectively. Lastly, safety concerns play a major role in construction, and the current maturity and reliability of AI technologies limit their practical application. To overcome these barriers, it is crucial to develop a responsible and trustworthy AI framework to guide the implementation of AI in construction.

It is worth noting that the existing research primarily focuses on the potential of AI technologies and the challenges of applying them to the construction industry [5,6]. However, the obstacles to innovation and progress are not limited to the construction industry itself but also the acceptance by various stakeholders and the unknown, uncontrollable risks that the technologies may introduce. Therefore, this paper, from the perspectives of AI impact and internal and external regulation, summarised the common characteristics of AI technologies and methods applicable to construction projects. It also proposed a general framework for designing, developing, and implementing AI in a reliable and trustworthy way. The main contributions of this paper include the following:

- Based on regulation documents published by international organisations and governments, we have identified and defined 13 principles for the design of an AI system to make it trustworthy and responsible.
- We proposed an operation framework to ensure the trustworthy and responsible implementation of AI throughout the entire system lifecycle in the construction industry.

This paper is organised as follows: Section II provides a literature review, covering (A) existing AI applications in the construction industry and (B) trustworthy and responsible AI frameworks that have been implemented in other industries, as well as relevant government regulations. Section III provides a detailed operation framework for ensuring trustworthy and responsible AI throughout its entire lifecycle from planning to archiving.

2. Literature Review

2.1. AI Applications in Construction

Over the past decade, with the rapid development of AI technologies in fields such as computer vision, natural language processing, robotics, and automated planning, more and more applications have been integrated into the construction industry across its lifecycle [6,7], including planning, design, and construction.

The planning phase involves addressing numerous factors such as project timelines, costs, supply chain logistics, risk management, environmental impact, and sustainability [8–10]. Eber [11] discussed the potential of using AI to break down complex construction projects into independent subsystems during the planning stage. Zuo et al. [12] developed a system using natural language processing to automatically retrieve risk cases for assessing risk management in construction projects. Hatami et al. [13] explored the potential for replacing traditional discrete event simulation with deep learning networks in construction project planning. Given the many influencing factors and the complexity of data structures in the planning phase, current research is still in the exploratory and discussion stages. There is a lack of mature and widely adopted systems or frameworks for the automation of construction project planning.

The design phase builds on the planning stage and is currently the most digitised part of the construction process. Commercial automation tools like AutoCAD are widely used in this phase, and Building Information Modelling (BIM) is well-developed to generate digital representations of the physical and functional characteristics of facilities and visualise the construction lifecycle [14]. Various data sources from the design phase generated a large amount of data. The utilisation of blockchain technology addressed problems related to data security, privacy, and modification tracking [15]. The combination of BIM and blockchain improved data auditing, provenance, and accountability within BIM systems. This approach enables data to be stored in a transparent, secure, and convenient environment [16], making it possible to generate a comprehensive and secure information and knowledge base in this phase [4]. However, a significant challenge remains in effectively incorporating AI to fully leverage these digital advancements and explore the full potential of this knowledge base.

The construction phase accounts for approximately 50% of the total carbon emissions over the entire lifecycle of a construction project [17], creating an urgent demand for AI solutions to help achieve sustainable development goals. This phase has seen a broad range of AI applications in recent years. Technologies such as computer vision, drone photogrammetry, and robotics have been employed to address issues like site safety, low resource utilisation, and project delays. This phase is the main focus of AI implementation in the construction industry.

Site safety encompasses aspects such as personal protective equipment (PPE) detection [18], hazard identification [19], unsafe behaviour monitoring [20], and dangerous zone detection [21]. The most common approach is using on-site cameras and employing computer vision and AI algorithms including object detection, tracking, and action recognition to identify PPE, misplaced equipment, or workers' behaviours and trajectories [5]. Integrating computer vision with drone technology, drone photogrammetry has been widely used for site monitoring and topography assessment [22]. Choi et al. [23] explored the potential of drone photogrammetry in BIM creation and digital site management. It discussed the impact of the drone flight parameters including altitude, camera angles, and distances during image capture on creating accurate point cloud models. Mihail et al. [24] used drones to capture the construction plant's trajectory, comparing optimal and working trajectories to estimate productivity, optimise mechanisation sequence, and reduce fuel consumption. The intelligent transformation of tasks such as material handling, assembly, and inspection in the construction phase not only relies on the support of AI algorithms but also requires the integration of robotics technology as a carrier [25]. Kayhani et al. [26] modelled a lifted object as a three-degree-of-freedom convex mobile robot and proposed a robotics path planning method to optimise and automate the lift path for heavy crawler cranes. Ali et al. [27] developed an automated robot facade location-finding tool to support activities such as facade and steel beam assembly. Additionally, AI technology-integrated robotics is able to inspect specific work conditions. Nguyen and La [28] developed a tank-like robot that can climb on flat steel surfaces to inspect in-depth fatigue cracks in steel structures. However, high operational costs and the inability to provide a stable, controlled work environment present challenge in the real application of robotics. These issues have resulted in slower adoption of robotics technology in construction compared to other industries such as manufacturing [29].

To sum up, most current AI applications in the construction sector are concentrated in the areas of perception (scene recognition) and action (robotic assistance). There is a lack of applications for decision optimisation, which has the potential for transforming low-productivity work patterns. Meanwhile, most existing areas of research remain at the stage of proof of concept and system implementation in specific scenarios. Due to factors such as technological maturity, controllability of the environment, and safety concerns, the acceptance of AI systems in on-site construction implementation is still relatively low. The currently developed AI systems have not yet achieved widespread, large-scale implementation. Enhancing the trustworthiness of AI systems applied in this field remains an unaddressed research gap.

2.2. Trustworthy and Responsible AI Frameworks

In recent years, various industries, including healthcare, transportation, and manufacturing, have begun their intelligent transformation. AI applications have significantly improved productivity [30,31]. However, the uncontrollable and unexpected risks associated with AI have raised public's concerns [32,33]. As a result, ensuring that AI systems are trustworthy and responsible has become an increasingly important research focus. Moreover, international organisations and governments have published guidelines for the development of trustworthy AI to address safety vulnerabilities and regulatory gaps arising from rapid technological advancements.

Regarding the term trustworthy and responsible AI, there is a lack of clear and explicit definitions in both academia and industry [34]. Most descriptions are based on the expected characteristics of such systems [35,36]. Moreover, different industries have diverse requirements and demands for AI applications. To achieve trustworthy and responsible AI, establishing clear design and development principles becomes particularly important. This paper reviews research on trustworthy AI across various industries and AI development guidelines published by international organisations and governments. Based on these insights, we summarise and define 13 principles for achieving trustworthy and responsible AI. These principles can be categorised into three parts, including internal characteristics of AI, interaction principles of AI, and social impact of AI, as illustrated in Figure 1.

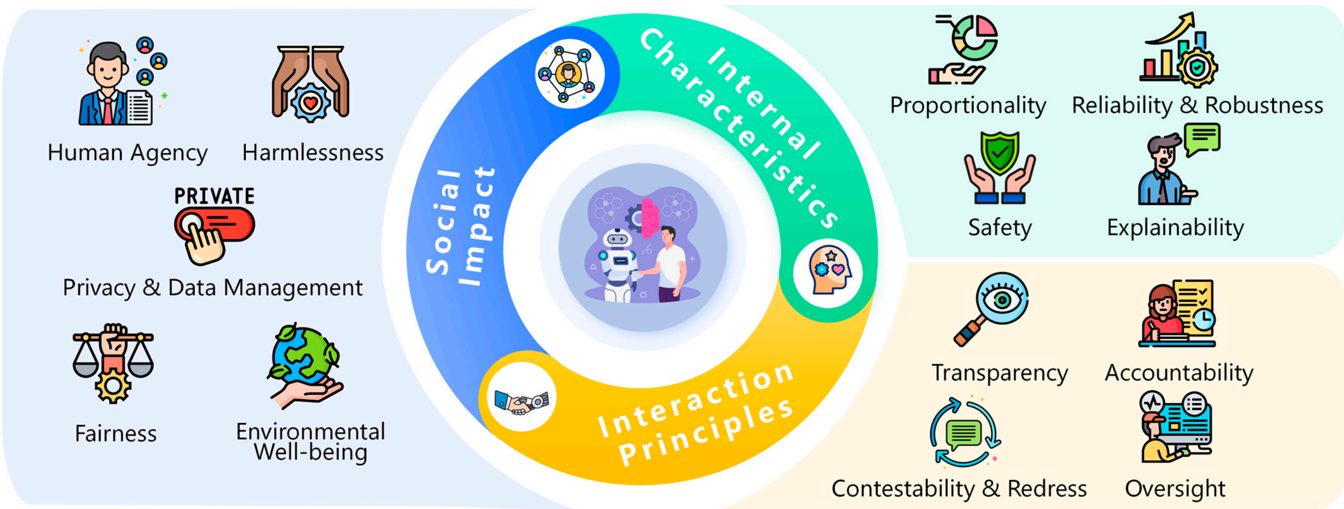

Figure 1. Principles to design trustworthy and responsible AI.

Internal characteristics are the technological characteristics that include the following:

- Proportionality (P): The AI system should have clearly defined purposes and applicable scopes. Any processes related to its lifecycle must not exceed what is necessary to achieve legitimate purposes or objectives.
- Reliability and Robustness (RR): The AI system should function reliably as intended and demonstrate robustness in unexpected situations.
- Safety (S): The system should be resilient against attempts to alter its use or performance by third parties and minimise unintended harm.
- Explainability (E): It should enable relevant parties to access, interpret, and understand its decision-making processes.

Interaction principles should be considered during the interaction between the AI system and all levels of stakeholders and participants and include the following:

- Transparency (T): The AI system should inform the relevant stakeholders of its capabilities and limitations and inform affected persons about their rights.
- Accountability (A): It should set clear lines of accountability across the life cycle.
- Contestability and Redress (CR): Participants in the AI life cycle should be able to contest an AI decision or outcome and provide feedback.

- Oversight (O): The supply and operation of the AI system should be appropriately controlled and overseen by humans.

Social impact is the social attributes that an AI system has that could affect individuals and communities. This includes the following:

- Human Agency (HA): The AI system should serve people while respecting human dignity and personal autonomy.
- Harmlessness (H): It should share the same values as humans, adhere to ethical standards, and prevent harms that can have material impact on people's lives.
- Privacy and Data Management (PD): The system should be used in accordance with privacy and data protection regulations.
- Fairness (F): It should promote equal access, gender equality, and cultural diversity while avoiding discriminatory impacts and unfair biases.
- Environmental Well-being (EW): The AI lifecycle should be sustainable and environmentally friendly, benefiting all human beings.

Table 1 presents the coverage of the defined principles in studies across different industry and guidance documents. It can be seen that most of the reviewed frameworks and government guidelines emphasise principles such as reliability, safety, explainability, transparency, privacy, and fairness, which are essential for enhancing public trust. However, in different sectors, more targeted principles and actions are adopted in designing trustworthy AI frameworks to address their unique characteristics and demands.

Table 1. Trustworthy and responsible AI principles that have been covered in the research and government documents.

Area	**P**	**RR**	**S**	**E**	**T**	**A**	**CR**	**O**	**HA**	**H**	**PD**	**F**	**EW**
Medical training [37]	✓	✓		✓	✓	✓		✓		✓	✓	✓	
Automated driving [38]	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		
Digital finance [39]		✓	✓	✓	✓			✓		✓	✓	✓	
B5G/6G [40]	✓	✓	✓	✓	✓			✓			✓	✓	✓
Air traffic control [41]		✓	✓	✓	✓		✓	✓		✓	✓	✓	
Military cyber operation [42]		✓	✓	✓	✓	✓		✓	✓	✓		✓	
IoT [43]		✓	✓								✓		✓
AR [44]		✓		✓	✓		✓		✓		✓		
Public institution [45]				✓	✓	✓	✓	✓			✓	✓	
Practice framework [46]		✓	✓	✓	✓	✓		✓			✓	✓	
Governance framework [47]		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	
Government	**P**	**RR**	**S**	**E**	**T**	**A**	**CR**	**O**	**HA**	**H**	**PD**	**F**	**EW**
UK regulation [48]		✓	✓	✓	✓	✓	✓	✓		✓		✓	
EU AI act [49]		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
USA framework [36]	✓	✓	✓	✓	✓	✓				✓	✓	✓	
Japan governance [50]		✓	✓	✓	✓	✓		✓	✓		✓	✓	✓
UNESCO recommendation [51]	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

In the healthcare sector, optimising decisions to improve patient health is crucial. Sqalli et al. [37] highlighted the importance of communication and collaboration among stakeholders in AI development and usage, including healthcare educators, clinicians, technical experts, and data scientists. Atakishiyev et al. [38] emphasised that effective user interaction design plays a critical role in autonomous driving for driving comfort and safety and suggested that AI systems should be designed from the user's perspective. Barmpounakis et al. [40] claimed that AI systems should possess real-time learning and anti-malicious attack capabilities in 6G communication, ensuring low latency to enhance resource efficiency and environmental friendliness. In air traffic controller operations, AI applications often operate in human–machine collaboration scenarios, requiring AI users to provide real-time feedback [41]. In public services, Laux et al. [45] explored incorporating stakeholder interaction and feedback mechanisms to enhance public trust. Li et al. [46] discussed the importance of the generalisation in AI systems in practice, reflecting their adaptability. Pursuing high generalisation often compromises accuracy, potentially reducing public acceptance of the technology. Therefore, at this stage, generalisation should not be a principle or goal for trustworthy and responsible AI. In digital finance, Kanaparthi [39] proposed that personalised financial services can enhance user trust, as reliable AI decisions in specific situations could increase user confidence. All these studies focus primarily on how to make AI more trustworthy from a technical perspective. In contrast, governments, when formulating guidelines and policies, concentrate more on the potential harms that AI technologies might cause to humanity, their impact on equality and sustainable development, and the need for more effective regulatory measures. Moreover, in the construction industry, especially for on-site construction, site safety is a top priority. The implemented AI systems should be harmless for construction safety. Also, human–AI collaboration is the most common scenario for the application of AI systems. Ensuring their reliability and protecting user privacy are crucial for increasing trust among end users.

3. Framework

Based on the design principles defined above, an operation framework for achieving trustworthy and responsible AI application in the construction industry is proposed and introduced in this section. This framework covers the focus on the whole system lifecycle and is presented in Figure 2. It categorises the process into six key phases: planning, data collection, algorithm development, deployment, maintenance, and archiving. The intermediate layer of the figure highlights the actions and focuses what needs to be done in each phase, while the outer layer presents corresponding principles.

3.1. Planning Phase

As the most strategically significant part, the planning phase determines the direction of the entire project [3].

3.1.1. Stakeholder Engagement

In the planning phase, it is crucial to understand the requirements of stakeholders and clearly define the targets. It requires close communication among all participants in the system development process, including developers, deployers, and end users, in line with the principle of transparency [52]. Such communication can effectively identify gaps between existing technologies and the expectations of end users, thereby helping set targets and determining the direction of development. Secondly, it is essential to define the evaluation metrics based on these targets [53]. These metrics should not only assess the performance of the AI methods but also evaluate the impact of their implementation and measure the overall success of the project. It provides clear guidance for subsequent phases of the process, reduces uncertainty during development, and enhances the system's explainability and accountability.

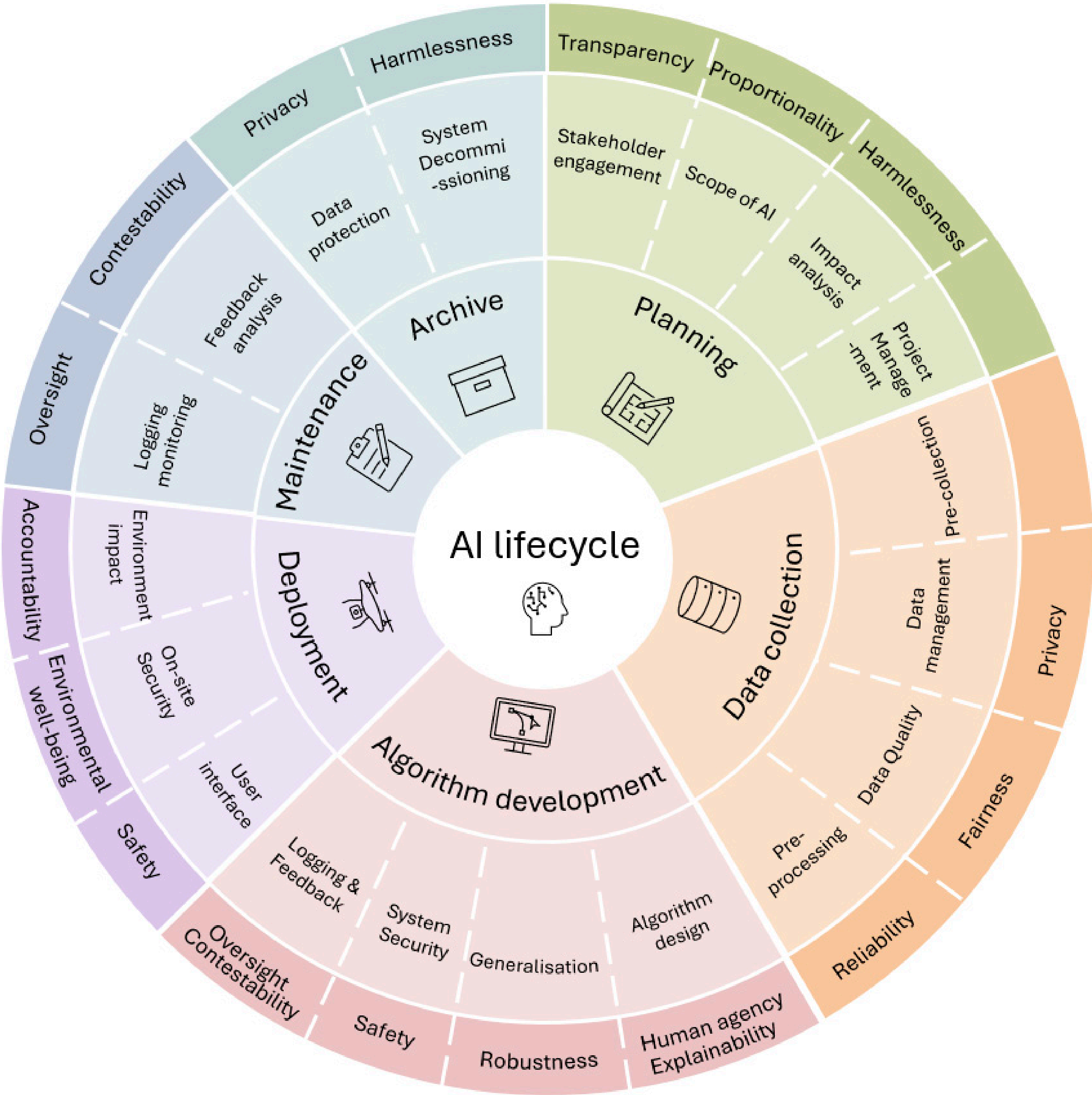

Figure 2. Operation framework for the entire AI lifecycle.

3.1.2. Scope of the System

Based on the principle of proportionality, defining the scope of the AI system is another critical task in the planning phase to ensure its trustworthiness and responsibility [49]. The system's capabilities must match the project's requirements, ensuring the technology's effectiveness while avoiding unnecessary complexity and cost overruns due to overdevelopment. It could reduce the barriers to implementing artificial intelligence applications in the construction industry [6]. The defined scope not only establishes the technical framework for the system but also significantly influences the allocation and management of resources.

3.1.3. Societal Impact

Given that the implementation of AI systems could have societal and environmental impacts and may raise issues related to ethics, privacy, and social equity [54], conducting forward-looking analyses and assessments of these potential impacts during the design and planning phase is vital to ensure the system's harmlessness and environmental friendliness.

3.1.4. Project Management

Given the significant uncertainty of AI development projects [55], implementing systematic project management and risk assessment helps ensure that the project proceeds as expected while minimising potential risks related to data and resource acquisition [56], technical implementation [57], and changes in the external environment [3]. Anticipating potential challenges during the planning phase and developing reasonable solutions to address them will contribute to the smooth operation of the entire AI lifecycle.

3.2. Data Collection Phase

For AI systems, the type and quality of data directly influence the algorithms' structure and model performance [58,59]. With the widespread use of big data, the data collection phase has become a complex and systematic procedure rather than a simple act of acquiring data. Several core actions involved in this phase are presented in Figure 2.

3.2.1. Pre-Collection

In the construction industry, commonly used data can be categorised as structured and unstructured types. Structured data are normally obtained with some manual definition and processing and are organised in the format of tables or constructed databases. Unstructured data come in various forms, including text, images, videos, audio, and specific formats from other sensors [60]. Before data collection, several things must be determined, including identifying variables, data types, and sensors. Identifying key variables is a process of characterising the problems the AI system aims to solve. By analysing these variables, the data types can be identified. Since the type, range, and precision of sensors used for data collection can directly affect performance and accuracy, the sensors or data collection tools should be properly selected in this phase.

3.2.2. Data Management

Designing effective data management and privacy protection methods is the key principle of trustworthy AI in the data collection process. Especially, for on-site construction AI applications, limitations in power and internet access could pose challenges for large-scale data collection, storage, and transmission. To address this, robust data management strategies must be implemented to ensure data integrity and traceability. As data volumes increase, particularly in the context of big data, solutions such as cloud computing and edge computing are widely applied in AI practice within the construction industry [61], which brings the potential for data leaks or attacks during acquisition, transmission, and storage [62]. Within the trustworthy AI framework, data security and privacy are critical aspects, particularly in applications involving sensitive user data (e.g., facial information). Therefore, it is important to ensure the security of data storage and transmission and protect personal privacy to avoid undue impacts in data collection and processing during the system operation. As a decentralised database technology, blockchain provides an unchangeable, distributed, and transparent approach to data sharing and storage [63]. The combination of AI and blockchain enables analytics and decision-making to be performed in a distributed and decentralised manner, avoiding single-point failure. The characteristic of immutability improves the AI model's resistance to data manipulation and accountability. The application of such a technology in AI-enabled excavator pose recognition has been proven to reduce the risk of being vulnerable to network attacks such as backdoor attacks and code tampering [64].

3.2.3. Data Quality

High-quality data are the foundation of reliable AI performance, and data shift is one of the primary causes of AI system failure while the working scenarios of the site for construction projects are normally diverse. To avoid introducing bias into the system, particular attention should be paid to the data source, the sample diversity, and whether the data fairly represent different groups and situations. The collected data should cover the various operation scenarios of the system, avoiding discriminatory or unfair outcomes [48,65]. Moreover, evaluating data quality before training can effectively control the development costs of AI systems in the construction industry. Outliers are observations that significantly deviate from other data samples in the dataset [66]. It can arise due to noise, inaccurate measurements, or data corruption during the data collection process for construction sites. The outliers can make an impact on the preference of the feature during the model training, and some malicious data can further compromise the model's performance. Therefore, removing outliers is crucial for maintaining data quality. Common outlier detection methods include statistical-based, distance-based, clustering-based, and deep learning-based approaches [67]. Another important part of data quality evaluation is the data distribution shift, which refers to the difference between the source domain and target domain in datasets. This is considered one of the main causes of AI unreliability in real-world applications. Data visualisation techniques are commonly employed to assess distribution shifts. Some data similarity metrics can be further used to quantify the distribution shift, such as KL-divergence and the Jensen–Shannon distance, which can be applied to evaluate the similarity between two probability distributions [68].

3.2.4. Data Preprocessing

Data preprocessing transforms raw data into a format suitable for algorithms to process. Properly designed preprocessing can effectively eliminate noise, reduce data redundancy, and ensure data are formatted and standardised, which significantly enhances model performance. Common preprocessing steps include data cleaning (handling missing data and outlier detection), data integration, data transformation (e.g., normalisation and standardisation), data reduction, and data splitting [69]. Depending on the application context, techniques such as data augmentation and feature engineering help ensure that the model has sufficient adaptability to handle complex and diverse scenarios. Furthermore, the growing use of generative AI and large language models (LLMs) require vast amounts of high-quality data. Preprocessing methods specifically for LLMs, such as filtering out harmful content, deduplication, privacy filtering, and tokenisation, are continuously being optimised to meet the unique requirements of these models [70].

3.3. Algorithm Development Phase

To ensure the system meets user requirements while remaining safe and reliable, the algorithm design must focus not only on its reliability and performance but also on factors such as safety, explainability, and social impact. The following parts need to be considered:

3.3.1. Algorithm Design

The algorithm should have a human-centred and needs-oriented design. The essence of AI is to serve humans, so algorithm development must be human-centred [49,71]. Developers should fully consider the real needs and expectations of users during the development phase, ensuring that the algorithm not only solves technical problems but also provides a good user experience with human–AI interactions. Explainability is one of the most critical factors determining the trustworthiness of AI systems [72]. For construction companies, which often purchase or license technology from external developers rather than develop it in-house, the opacity of decisions made by AI-based systems can make stakeholders more hesitant to trust them [73,74]. In the design phase, modular design is an important strategy to enhance the explainability and accountability of AI systems. By decomposing the algorithm into independent functional modules, developers and users can more easily understand how the model works, evaluate the performance of individual modules, and clarify accountability mechanisms. Modular design also simplifies testing, maintenance, and upgrades, which improves system flexibility and robustness. This design approach is particularly well-suited for industries like construction, where high explainability and clear accountability are essential.

3.3.2. Model Generalisation

Generalisation is one of the most critical objectives in AI algorithm design. A model's generalisation capability reflects its adaptability to unseen data, allowing it to effectively handle the diversity and complexity of real-world scenarios. Common strategies to avoid overfitting, such as proper dataset partitioning and cross-validation, regularisation techniques, and data augmentation, can significantly enhance a model's generalisation ability. In the construction industry, on-site digitisation is limited, and data scarcity remains a significant challenge, particularly for camera-based systems installed on construction equipment or facilities. As construction progresses, changes in the environment, site layout, and camera locations further impact data quality, posing additional challenges for these systems [56]. Therefore, it is essential to design models that can adapt to diverse and complex scenarios with limited data. AI paradigms such as transfer learning, contrastive learning, and few-shot learning offer great potential in addressing these challenges [75,76].

3.3.3. Misuse and System Security

Potential risks for misuse need to be considered in the development phase. The capabilities of an AI model must be constrained within reasonable limits to prevent unintended or harmful consequences [48]. For instance, computer vision-based methods developed for worker safety monitoring and PPE detection on construction sites could also be used to analyse worker's performance [77,78]. However, the misuse of such surveillance technology could potentially violate privacy or enable unethical monitoring. To mitigate such risks, developers should implement use restrictions tailored to the application context to prevent misuse. In terms of cyber security, the complexity of AI systems makes them potential targets for attacks, particularly in public network environments. To prevent malicious attacks, data breaches, and system disruptions, developers must incorporate necessary security measures into the algorithm design. These include defences against adversarial attacks, encryption technologies, and access control policies [79,80]. These measures could effectively reduce security risks and ensure the safe operation of AI systems in construction environments.

3.3.4. Logging and Feedback System

A well-developed AI system must include comprehensive logging capabilities. Logs allow detailed tracking of system usage, providing insights into the model's operational status, performance, and any encountered anomalies [49]. This not only provides developers with evidence for system maintenance but also offers explainability and accountability for regulatory authorities, ensuring the system is used legally and compliantly. Designing a feedback system can significantly enhance the efficiency of AI system maintenance and updates. User feedback helps developers identify issues and deficiencies in real-world applications, guiding subsequent optimisation and upgrades [48]. A well-designed feedback system can quickly capture user pain points, facilitating iterative updates and continuous improvement of the algorithm.

3.4. Deployment Phase

The lack of skilled technicians and concerns about potential violations of personal privacy when working alongside AI systems are the main barriers to intelligent transformation in the construction phase. Therefore, to make AI systems more trustworthy and acceptable, in the deployment phase, the design of AI systems should focus on three key principles: transparency, security, and environmental impact.

3.4.1. End-User Interaction

Designing an efficient and clear user interaction approach is essential during the AI system deployment phase. Simple and straightforward instructions can ensure that end users on the construction site operate the system correctly and efficiently. At this stage, the system should provide comprehensive user guides and feature explanations, along with a user-friendly interface to enhance usability. For robotics and drone-based automated systems used on construction sites that require worker operation and supervision, such as robotic arms for wall construction [81], mobile robotic systems for brickwork [82], and surveying and mapping drone systems [22], end users should be educated and have a clearly understanding of how the system operates and how data are stored and processed in order to alleviate concerns and avoid misuse. Moreover, clarifying accountability mechanisms during deployment is crucial for preventing misuse or abuse. This ensures that all participants involved in the operation of the AI system (developers, deployment personnel, users, etc.) are aware of their responsibilities and can rectify inappropriate actions if necessary.

3.4.2. On-Site Security

A major challenge in deploying AI systems in the construction industry is ensuring the security of hardware devices. Sensors and other equipment are often exposed to outdoor environments for extended periods, making them susceptible to physical damage, harsh weather conditions, and even vandalism. To ensure the long-term and stable operation of these devices, appropriate security and protective measures must be implemented during deployment. For example, using durable, waterproof hardware casings, employing remote monitoring technologies to oversee equipment status, and establishing regular maintenance procedures can reduce the risk of hardware failure. Additionally, ensuring that the deployment of AI hardware does not interfere with normal construction activities is critical for on-site applications. When deploying AI sensors on construction sites, factors like equipment placement, wiring methods, and safety must be considered to avoid disrupting the efficiency and safety of construction workers and machinery. Moreover, for large robotic systems such as autonomous walking excavators, proper planning of the working area and collaboration mechanisms with workers become increasing important [83]. Thus, prior to deployment, it is essential to maintain open communication with the construction team and develop a reasonable deployment plan to ensure hardware operates effectively without disrupting ongoing construction activities.

3.4.3. Environmental Impact

In construction, AI system deployment often involves not only software but also hardware installations, such as sensors, cameras, and drones. The deployment of this equipment can potentially disrupt wildlife habitats near construction sites and generate noise pollution [84]. Therefore, conducting an environmental impact assessment during deployment is necessary to ensure that equipment deployment does not cause irreversible damage to the natural environment. It should also consider indirect impacts, such as energy consumption. AI systems typically require significant amounts of energy to operate [85]. A feasible solution to address this challenge is the use of energy-efficient equipment and the implementation of intelligent management systems to reduce energy consumption.

3.5. Maintenance Phase

The trust in the AI system is closely tied to the continuous and reliable outputs generated by the system. This is particularly important in construction projects, where complex and dynamic environments bring the challenge to the reliability and robustness of AI systems. When systems make wrong decisions or provide unsafe instructions, user trust can be significantly diminished [74]. Therefore, ensuring the continuous functionality of AI systems during the maintenance phase and allowing the system to evolve as needs change are key tasks.

3.5.1. Daily Performance Monitoring and Failure Analysis

Continuous monitoring of an AI system's daily performance is critical for maintaining reliability and consistency. This involves tracking both the performance of the model and the hardware and software components. Regular evaluations enable developers to detect potential issues in the early stage. Causes of failure might include data drift, environment changes, hardware malfunction, or external security threats [86]. Once issues are identified, corrective measures should be promptly taken to prevent negative impacts on users. For instance, in a construction site, the performance of a computer vision-based AI system could be affected by changes in the environment, lighting, and objects within the scene [56]. Regularly assessing model performance and making necessary adjustments are essential to maintaining the reliability of the system.

3.5.2. Handling User Feedback and System Updates

In the maintenance phase, user feedback helps identify shortcomings in the system's operation in various real-world scenarios. Analysing feedback can reveal unexpected problems in the system and help guide improvement strategies. Particularly in construction projects, user requirements may evolve as the project progresses. Actively responding to feedback and changing needs, while continuously updating the system, is key to ensuring its robustness and trustworthiness. System updates may involve algorithm optimisation, model updates, user interface improvements, or hardware upgrades. It should be noted that all updates must be rigorously tested and validated to ensure that the system remains stable and efficient post-deployment.

3.6. Archive Phase

The archive phase is the end of the AI lifecycle and ensures secure termination of the system without risking misuse. Key aspects for the archive phase include the following:

3.6.1. Data Protection

AI systems handle vast amounts of sensitive data, such as user information, environmental data, and business records. Robust data protection methods must be taken to prevent data leaks or unauthorised use. This includes encryption, access controls, and securely deleting or archiving data to ensure unauthorised parties cannot access it [87]. The use of the data should comply with legal frameworks, such as the General Data Protection Regulation (GDPR) [88]. Developers must ensure that all data retention and archive processes comply with these regulations, especially in cases involving personal privacy. In the construction industry, the lack of high-quality data and the versatility of the data that can be used for different projects mean that data can often be reused. Ensuring that data are safely stored in a format that supports reusability is an important consideration during the archiving phase.

3.6.2. System Decommissioning and Preventing Misuse

Proper decommissioning of the AI system itself is equally important when it reaches the end of its lifecycle [48]. The system's code, models, and hardware need to be handled responsibly to prevent misuse or unauthorised access. Hardware devices must be thoroughly wiped of data, and steps should be taken to ensure they cannot be repurposed without authorisation. Similarly, the AI models should be carefully managed. In some cases, it may be prudent to destroy the models or store them in highly controlled environments to prevent them from being used in inappropriate or non-compliant scenarios. Both the maintenance and archive phases are fundamental to maintaining the security, performance, and compliance of AI systems.

4. Conclusions

Enhancing the trustworthiness of AI systems is essential for their acceptance and widespread adoption, especially in the construction industry, which is significantly behind other sectors in terms of digital transformation. This paper reviewed existing frameworks for trustworthy and responsible AI across various fields, as well as guidelines and regulation issued by governments. We proposed 13 principles for the whole life cycle of a trustworthy AI system and categorised them into three dimensions: internal characteristics, interaction principles, and social impact. Based on the proposed principles and the factors or concerns that hinder the implementation of AI systems in the construction industry, such as decision opacity, on-site safety, and data privacy, an operation framework has been proposed for the entire AI lifecycle, specifically tailored to applications in construction. This framework integrates AI design principles and highlights critical areas of focus at different stages of AI development to enhance trust between humans and AI systems in construction environments. The developed framework can serve as a guideline for the development of on-site AI systems, improve trust and acceptance, and promote their widespread use.

It should be noted that this proposed framework was developed from a general AI system development perspective and tailored specifically for implementation in on-site construction works in this industry. However, other AI applications used in the planning, design, and maintenance phases of the construction and building lifecycle have not been fully explored in this paper. For instance, with the development of large language models, attempts have been made in the construction industry to enable information retrieval from BIM [89], compliance checking with legal requirements [90], and construction task scheduling [91]. Although the development process of such AI systems follows the proposed framework, the focus on trustworthy and responsible AI is slightly different. Principles such as proportionality, oversight, harmlessness, and human agency are much more important. Moreover, this framework provides guidance on actions required during the development phase. However, practical implementation approaches for different technologies could vary depending on specific tasks, which still requires further exploration. In future works, a more comprehensive framework that covers all phases of the building lifecycle should be established based on the existing one.

Author Contributions: Conceptualization, L.Y. and Y.Z.; methodology, L.Y.; software, L.Y.; validation, L.Y., Y.Z. and Z.Z.; formal analysis, L.Y.; investigation, L.Y.; resources, L.Y. and Z.Z.; data curation, L.Y.; writing—original draft preparation, L.Y.; writing—review and editing, Y.Z., Z.Z. and G.A.; visualization, L.Y. and Z.Z.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z. and G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by "TRAMS-Enterprise—Trustworthy, Responsible AI and ML for construction, Secure and Enterprise Ready" funded by Innovate UK (Ref: 10093011).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. United Nations Department of Economic and Social Affairs. *Population Division (2024). World Population Prospects 2024: Summary of Results*; United Nations Department of Economic and Social Affairs: New York, NY, USA, 2024.
2. United Nations Environment Programme and Yale Center for Ecosystems + Architecture. *Building Materials and the Climate: Constructing a New Future*; United Nations Environment Programme and Yale Center for Ecosystems + Architecture: Yale, CT, USA, 2023.
3. Regona, M.; Yigitcanlar, T.; Xia, B.; Li, R.Y.M. Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 45. [CrossRef]
4. Datta, S.D.; Islam, M.; Rahman Sobuz, M.H.; Ahmed, S.; Kar, M. Artificial Intelligence and Machine Learning Applications in the Project Lifecycle of the Construction Industry: A Comprehensive Review. *Heliyon* **2024**, *10*, e26888. [CrossRef] [PubMed]
5. Rabbi, A.B.K.; Jeelani, I. AI Integration in Construction Safety: Current State, Challenges, and Future Opportunities in Text, Vision, and Audio Based Applications. *Autom. Constr.* **2024**, *164*, 105443. [CrossRef]
6. Abioye, S.O.; Oyedele, L.O.; Akanbi, L.; Ajayi, A.; Davila Delgado, J.M.; Bilal, M.; Akinade, O.O.; Ahmed, A. Artificial Intelligence in the Construction Industry: A Review of Present Status, Opportunities and Future Challenges. *J. Build. Eng.* **2021**, *44*, 103299. [CrossRef]
7. Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep Learning in the Construction Industry: A Review of Present Status and Future Innovations. *J. Build. Eng.* **2020**, *32*, 101827. [CrossRef]
8. Han, Y.; Yan, X.; Piroozfar, P. An Overall Review of Research on Prefabricated Construction Supply Chain Management. *Eng. Constr. Archit. Manag.* **2023**, *30*, 5160–5195. [CrossRef]
9. Yin, D.; Chen, Y.; Jia, H.; Wang, Q.; Chen, Z.; Xu, C.; Li, Q.; Wang, W.; Yang, Y.; Fu, G.; et al. Sponge City Practice in China: A Review of Construction, Assessment, Operational and Maintenance. *J. Clean. Prod.* **2021**, *280*, 124963. [CrossRef]
10. Zhao, X. Construction Risk Management Research: Intellectual Structure and Emerging Themes. *Int. J. Constr. Manag.* **2024**, *24*, 540–550. [CrossRef]
11. Eber, W. Potentials of Artificial Intelligence in Construction Management. *Organ. Technol. Manag. Constr.* **2020**, *12*, 2053–2063. [CrossRef]
12. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving Similar Cases for Construction Project Risk Management Using Natural Language Processing Techniques. *Autom. Constr.* **2017**, *80*, 66–76. [CrossRef]
13. Hatami, M.; Franz, B.; Paneru, S.; Flood, I. Using Deep Learning Artificial Intelligence to Improve Foresight Method in the Optimization of Planning and Scheduling of Construction Processes. In Proceedings of the Computing in Civil Engineering 2021—Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021, Orlando, FL, USA, 12–14 September 2021.
14. Wu, Z.; Chen, C.; Cai, Y.; Lu, C.; Wang, H.; Yu, T. BIM-Based Visualization Research in the Construction Industry: A Network Analysis. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3473. [CrossRef] [PubMed]
15. Pan, Y.; Zhang, L. Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions. *Arch. Comput. Methods Eng.* **2023**, *30*, 1081–1110. [CrossRef]
16. Hunhevicz, J.J.; Hall, D.M. Do You Need a Blockchain in Construction? Use Case Categories and Decision Framework for DLT Design Options. *Adv. Eng. Inform.* **2020**, *45*, 101094. [CrossRef]
17. Chen, L.; Huang, L.; Hua, J.; Chen, Z.; Wei, L.; Osman, A.I.; Fawzy, S.; Rooney, D.W.; Dong, L.; Yap, P.S. Green Construction for Low-Carbon Cities: A Review. *Environ. Chem. Lett.* **2023**, *21*, 1627–1657. [CrossRef]
18. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef]
19. Jeelani, I.; Asadi, K.; Ramshankar, H.; Han, K.; Albert, A. Real-Time Vision-Based Worker Localization & Hazard Detection for Construction. *Autom. Constr.* **2021**, *121*, 103448. [CrossRef]
20. Hung, P.D.; Su, N.T. Unsafe Construction Behavior Classification Using Deep Convolutional Neural Network. *Pattern Recognit. Image Anal.* **2021**, *31*, 271–284. [CrossRef]
21. Huang, H.; Hu, H.; Xu, F.; Zhang, Z.; Tao, Y. Skeleton-Based Automatic Assessment and Prediction of Intrusion Risk in Construction Hazardous Areas. *Saf. Sci.* **2023**, *164*, 106150. [CrossRef]
22. Choi, H.W.; Kim, H.J.; Kim, S.K.; Na, W.S. An Overview of Drone Applications in the Construction Industry. *Drones* **2023**, *7*, 515. [CrossRef]

23. Choi, W.; Na, S.; Heo, S. Integrating Drone Imagery and AI for Improved Construction Site Management through Building Information Modeling. *Buildings* **2024**, *14*, 1106. [CrossRef]
24. Mihail, S.I.; Oana, T.; Magdalena, C.; Radu, P.L.; Cristina, S.-G.; Ruxandra, S.A.; Alexandros, M. Aspects of the Use of Drone Photogrammetry in Construction Equipment Management Using Artificial Intelligence AI. *Rom. J. Transp. Infrastruct.* **2023**, *12*, 1–18. [CrossRef]
25. Xiao, B.; Chen, C.; Yin, X. Recent Advancements of Robotics in Construction. *Autom. Constr.* **2022**, *144*, 104591. [CrossRef]
26. Kayhani, N.; Taghaddos, H.; Mousaei, A.; Behzadipour, S.; Hermann, U. Heavy Mobile Crane Lift Path Planning in Congested Modular Industrial Plants Using a Robotics Approach. *Autom. Constr.* **2021**, *122*, 103508. [CrossRef]
27. Ali, A.K.; Lee, O.J.; Song, H. Robot-Based Facade Spatial Assembly Optimization. *J. Build. Eng.* **2021**, *33*, 101556. [CrossRef]
28. Nguyen, S.T.; La, H.M. A Climbing Robot for Steel Bridge Inspection. *J. Intell. Robot. Syst. Theory Appl.* **2021**, *102*, 1–21. [CrossRef]
29. Attalla, A.; Attalla, O.; Moussa, A.; Shafique, D.; Raean, S.B.; Hegazy, T. Construction Robotics: Review of Intelligent Features. *Int. J. Intell. Robot. Appl.* **2023**, *7*, 535–555. [CrossRef]
30. Alowais, S.A.; Alghamdi, S.S.; Alsuhebany, N.; Alqahtani, T.; Alshaya, A.I.; Almohareb, S.N.; Aldairem, A.; Alrashed, M.; Bin Saleh, K.; Badreldin, H.A.; et al. Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *BMC Med. Educ.* **2023**, *23*, 1–15. [CrossRef]
31. Iyer, L.S. AI Enabled Applications towards Intelligent Transportation. *Transp. Eng.* **2021**, *5*, 100083. [CrossRef]
32. Sharma, S. Benefits or Concerns of AI: A Multistakeholder Responsibility. *Futures* **2024**, *157*, 103328. [CrossRef]
33. Hendrycks, D.; Mazeika, M.; Woodside, T. An Overview of Catastrophic AI Risks. *arXiv* **2023**, arXiv:2306.12001.
34. Dehghani, F.; Dibaji, M.; Anzum, F.; Dey, L.; Basdemir, A.; Bayat, S.; Boucher, J.-C.; Drew, S.; Eaton, S.E.; Frayne, R.; et al. Trustworthy and Responsible AI for Human-Centric Autonomous Decision-Making Systems. *arXiv* **2024**, arXiv:2408.15550.
35. OECD. *Tools for Trustworthy AI*; OECD Digital Economy Papers; OECD: Paris, France, 2021; Volume 312.
36. Tabassi, E. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*; NIST: Gaithersburg, MD, USA, 2023.
37. Tahri Sqalli, M.; Aslonov, B.; Gafurov, M.; Nurmatov, S. Humanizing AI in Medical Training: Ethical Framework for Responsible Design. *Front. Artif. Intell.* **2023**, *6*, 1189914. [CrossRef] [PubMed]
38. Atakishiyev, S.; Salameh, M.; Yao, H.; Goebel, R. Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions. *IEEE Access* **2024**, *12*, 101603–101625. [CrossRef]
39. Kanaparthi, V. AI-Based Personalization and Trust in Digital Finance. *arXiv* **2024**, arXiv:2401.15700.
40. Barmpounakis, S.; Demestichas, P. Framework for Trustworthy AI/ML in B5G/6G. In Proceedings of the 2022 1st International Conference on 6G Networking, 6GNet 2022, Paris, France, 6–8 July 2022.
41. Stefani, T.; Deligiannaki, F.; Berro, C.; Jameel, M.; Hunger, R.; Bruder, C.; Kruger, T. Applying the Assessment List for Trustworthy Artificial Intelligence on the Development of AI Supported Air Traffic Controller Operations. In Proceedings of the AIAA/IEEE Digital Avionics Systems Conference—Proceedings, Barcelona, Spain, 1–5 October 2023.
42. Maathuis, C. Human Centered Explainable AI Framework for Military Cyber Operations. In Proceedings of the MILCOM 2023–2023 IEEE Military Communications Conference: Communications Supporting Military Operations in a Contested Environment, Boston, MA, USA, 30 October–3 November 2023.
43. Zhang, Q.; Zhong, H.; Shi, W.; Liu, L. A Trusted and Collaborative Framework for Deep Learning in IoT. *Comput. Netw.* **2021**, *193*, 108055. [CrossRef]
44. Xu, X.; Yu, A.; Jonker, T.R.; Todi, K.; Lu, F.; Qian, X.; Evangelista Belo, J.M.; Wang, T.; Li, M.; Mun, A.; et al. XAIR: A Framework of Explainable AI in Augmented Reality. In Proceedings of the Conference on Human Factors in Computing Systems—Proceedings, Hamburg, Germany, 23–28 April 2023.
45. Laux, J.; Wachter, S.; Mittelstadt, B. Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk. *Regul. Gov.* **2024**, *18*, 3–32. [CrossRef]
46. Li, B.; Qi, P.; Liu, B.; Di, S.; Liu, J.; Pei, J.; Yi, J.; Zhou, B. Trustworthy AI: From Principles to Practices. *ACM Comput. Surv.* **2023**, *55*, 1–46. [CrossRef]
47. Sharma, S. Trustworthy Artificial Intelligence: Design of AI Governance Framework. *Strateg. Anal.* **2023**, *47*, 443–464. [CrossRef]
48. Department for Science Innovation and Technology. *A Pro-Innovation Approach to AI Regulation*; Dandy Booksellers Ltd.: London, UK, 2023.
49. European Parliament The Artificial Intelligence Act of the EU. *Off. J. Eur. Union.* **2024**. Available online: https://artificialintelligenceact.eu/ai-act-explorer/ (accessed on 20 December 2024).
50. Expert Group on How Ai Principles Should Be Implemented. *AI Governance in Japan Ver. 1.1*; Japan's Ministry of Economy, Trade and Industry: Tokyo, Japan, 2021.
51. United Nations Educational, Scientific and Cultural Organization. *Recommendation on the Ethics of Artificial Intelligence*; United Nations Educational, Scientific and Cultural Organization: Paris, France, 2022.
52. Lucic, A.; Srikumar, M.; Bhatt, U.; Xiang, A.; Taly, A.; Liao, Q.V.; de Rijke, M. A Multistakeholder Approach Towards Evaluating AI Transparency Mechanisms. *arXiv* **2021**, arXiv:2103.14976.

53. Thomas, R.L.; Uminsky, D. Reliance on Metrics Is a Fundamental Challenge for AI. *Patterns* **2022**, *3*, 100476. [CrossRef] [PubMed]
54. Qian, Y.; Siau, K.L.; Nah, F.F. Societal Impacts of Artificial Intelligence: Ethical, Legal, and Governance Issues. *Soc. Impacts* **2024**, *3*, 100040. [CrossRef]
55. Vial, G.; Cameron, A.F.; Giannelia, T.; Jiang, J. Managing Artificial Intelligence Projects: Key Insights from an AI Consulting Firm. *Inf. Syst. J.* **2023**, *33*, 669–691. [CrossRef]
56. Zhang, B.; Yang, B.; Wang, C.; Wang, Z.; Liu, B.; Fang, T. Computer Vision-Based Construction Process Sensing for Cyber–Physical Systems: A Review. *Sensors* **2021**, *21*, 5468. [CrossRef]
57. Daniel, E.I.; Oshodi, O.S.; Nwankwo, N.; Emuze, F.A.; Chinyio, E. Barriers to the Application of Digital Technologies in Construction Health and Safety: A Systematic Review. *Buildings* **2024**, *14*, 2386. [CrossRef]
58. Whang, S.E.; Roh, Y.; Song, H.; Lee, J.-G. Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective. *VLDB J.* **2023**, *32*, 791–813. [CrossRef]
59. Schwabe, D.; Becker, K.; Seyferth, M.; Klaß, A.; Schaeffter, T. The METRIC-Framework for Assessing Data Quality for Trustworthy AI in Medicine: A Systematic Review. *npj Digit. Med.* **2024**, *7*, 203. [CrossRef]
60. Li, F.; Laili, Y.; Chen, X.; Lou, Y.; Wang, C.; Yang, H.; Gao, X.; Han, H. Towards Big Data Driven Construction Industry. *J. Ind. Inf. Integr.* **2023**, *35*, 100483. [CrossRef]
61. Reaño, C.; Riera, J.V.; Romero, V.; Morillo, P.; Casas-Yrurzum, S. A Cloud-Edge Computing Architecture for Monitoring Protective Equipment. *J. Cloud Comput.* **2024**, *13*, 82. [CrossRef]
62. Xiao, X.; Chen, C.; Skitmore, M.; Li, H.; Deng, Y. Exploring Edge Computing for Sustainable CV-Based Worker Detection in Construction Site Monitoring: Performance and Feasibility Analysis. *Buildings* **2024**, *14*, 2299. [CrossRef]
63. Tao, X.; Das, M.; Liu, Y.; Cheng, J.C.P. Distributed Common Data Environment Using Blockchain and Interplanetary File System for Secure BIM-Based Collaborative Design. *Autom. Constr.* **2021**, *130*, 103851. [CrossRef]
64. Zheng, C.; Tao, X.; Dong, L.; Zukaib, U.; Tang, J.; Zhou, H.; Cheng, J.C.P.; Cui, X.; Shen, Z. Decentralized Artificial Intelligence in Construction Using Blockchain. *Autom. Constr.* **2024**, *166*, 105669. [CrossRef]
65. Tawakuli, A.; Engel, T. Make Your Data Fair: A Survey of Data Preprocessing Techniques That Address Biases in Data towards Fair AI. *J. Eng. Res.* 2024, in press. [CrossRef]
66. Breuniq, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying Density-Based Local Outliers. *ACM SIGMOD Rec.* **2000**, *29*, 93–104. [CrossRef]
67. Smiti, A. A Critical Overview of Outlier Detection Methods. *Comput. Sci. Rev.* **2020**, *38*, 100306. [CrossRef]
68. Flovik, V. Quantifying Distribution Shifts and Uncertainties for Enhanced Model Robustness in Machine Learning Applications. *arXiv* **2024**, arXiv:2405.01978.
69. Maharana, K.; Mondal, S.; Nemade, B. A Review: Data Pre-Processing and Data Augmentation Techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [CrossRef]
70. Zhang, H.; Dong, Y.; Xiao, C.; Oyamada, M. Large Language Models as Data Preprocessors. *arXiv* **2023**, arXiv:2308.16361.
71. Tahaei, M.; Constantinides, M.; Quercia, D.; Muller, M. A Systematic Literature Review of Human-Centered, Ethical, and Responsible AI. *arXiv* **2023**, arXiv:2302.05284.
72. Pavlin, G.; De Villiers, J.P.; Ziegler, J.; Jousselme, A.L.; Costa, P.; Laskey, K.; De Waal, A.; Blasch, E.; Jansen, L. Relations between Explainability, Evaluation and Trust in AI-Based Information Fusion Systems. In Proceedings of the 2021 IEEE 24th International Conference on Information Fusion, Sun City, South Africa, 1–4 November 2021.
73. Markus, A.F.; Kors, J.A.; Rijnbeek, P.R. The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies. *J. Biomed. Inform.* **2021**, *113*, 103655. [CrossRef]
74. Emaminejad, N.; Akhavian, R. Trustworthy AI and Robotics: Implications for the AEC Industry. *Autom. Constr.* **2022**, *139*, 104298. [CrossRef]
75. Le-Khac, P.H.; Healy, G.; Smeaton, A.F. Contrastive Representation Learning: A Framework and Review. *IEEE Access* **2020**, *8*, 193907–193934. [CrossRef]
76. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A Comprehensive Survey of Few-Shot Learning: Evolution, Applications, Challenges, and Opportunities. *ACM Comput. Surv.* **2023**, *55*, 1–40. [CrossRef]
77. Hou, X.; Li, C.; Fang, Q. Computer Vision-Based Safety Risk Computing and Visualization on Construction Sites. *Autom. Constr.* **2023**, *156*, 105129. [CrossRef]
78. Li, J.; Miao, Q.; Zou, Z.; Gao, H.; Zhang, L.; Li, Z.; Wang, N. A Review of Computer Vision-Based Monitoring Approaches for Construction Workers' Work-Related Behaviors. *IEEE Access* **2024**, *12*, 7134–7155. [CrossRef]
79. Baniecki, H.; Biecek, P. Adversarial Attacks and Defenses in Explainable Artificial Intelligence: A Survey. *Inf. Fusion.* **2024**, *107*, 102303. [CrossRef]
80. Upreti, R.; Lind, P.G.; Elmokashfi, A.; Yazidi, A. Trustworthy Machine Learning in the Context of Security and Privacy. *Int. J. Inf. Secur.* **2024**, *23*, 2287–2314. [CrossRef]

81. Więckowski, A. "JA-WA"—A Wall Construction System Using Unilateral Material Application with a Mobile Robot. *Autom. Constr.* **2017**, *83*, 19–28. [CrossRef]
82. Dörfler, K.; Sandy, T.; Giftthaler, M.; Gramazio, F.; Kohler, M.; Buchli, J. Mobile Robotic Brickwork. In *Robotic Fabrication in Architecture, Art. and Design 2016*; Springer: Berlin/Heidelberg, Germany, 2016.
83. Jud, D.; Kerscher, S.; Wermelinger, M.; Jelavic, E.; Egli, P.; Leemann, P.; Hottiger, G.; Hutter, M. HEAP—The Autonomous Walking Excavator. *Autom. Constr.* **2021**, *129*, 103783. [CrossRef]
84. Koemle, D.; Zinngrebe, Y.; Yu, X. Highway Construction and Wildlife Populations: Evidence from Austria. *Land. Use Policy* **2018**, *73*, 447–457. [CrossRef]
85. Sangarya, V.; Bradford, R.; Kim, J.-E. Estimating Environmental Cost Throughout Model's Adaptive Life Cycle. *arXiv* **2024**, arXiv:2408.01446. [CrossRef]
86. Fang, W.; Ding, L.; Love, P.E.D.; Luo, H.; Li, H.; Peña-Mora, F.; Zhong, B.; Zhou, C. Computer Vision Applications in Construction Safety Assurance. *Autom. Constr.* **2020**, *110*, 103013. [CrossRef]
87. Colavizza, G.; Blanke, T.; Jeurgens, C.; Noordegraaf, J. Archives and AI: An Overview of Current Debates and Future Perspectives. *J. Comput. Cult. Herit.* **2022**, *15*, 1–15. [CrossRef]
88. Sartor, G.; Lagioia, F. *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence*; European Parliament: Brussels, Belgium, 2020. [CrossRef]
89. Zheng, J.; Fischer, M. BIM-GPT: A Prompt-Based Virtual Assistant Framework for BIM Information Retrieval. *arXiv* **2023**, arXiv:2304.09333.
90. Fuchs, S.; Witbrock, M.; Dimyadi, J.; Amor, R. Using Large Language Models for the Interpretation of Building Regulations. *arXiv* **2024**, arXiv:2407.21060.
91. Prieto, S.A.; Mengiste, E.T.; García de Soto, B. Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings* **2023**, *13*, 857. [CrossRef]

Achieving on-site trustworthy AI implementation in the construction industry: a framework across the AI lifecycle

Yang, Lichao

2024-12-25

Yang L, Allen G, Zhang Z, Zhao Y. (2024) Achieving on-site trustworthy AI implementation in the construction industry: a framework across the AI lifecycle. Buildings, Volume 15, Issue 1, January 2025, Article number 21

https://doi.org/10.3390/buildings15010021